STAT 211

How to do grouping?

Guidelines for Constructing a Grouped Frequency Distribution

1. each data item must be in one and only one class

2. must accommodate the largest and smallest item

3. number of classes:
   - if too many classes – not much improvement over raw data
   - if too few classes – we lose much detail, lost special features

4. table is easier to read if we use multiples of say 5, or 10, or 25, 100, 1000, …

Grouped Frequency Distribution (cont'd)

5. if possible, use equal class widths

6. avoid open classes, if possible

7. **class intervals and/or ranges should be given to the same degree of accuracy as the raw data**

8. all tables should have titles, and all columns should have headings (with units)

Advantages of a Grouped Frequency Distribution

- presents data in compact form

- shows overall picture

- the general nature of the data can be seen that otherwise may be lost in the detail of the raw data

Some things are lost. For example, the order of the data, and the exact data itself are lost.

A **Histogram** is a graph that consists of a series of rectangles such that:

1. the center of the base of each rectangle is at the class midpoint – the edges of the rectangles are at the class boundaries

2. the area is proportional to the frequency

To calculate the height of a block over a class interval :

❖ <u>divide</u> the <u>percent</u> by the length of the interval, ie. by the number of common units in that interval

if the table does not show percent, change the frequency to percent